

## RESEARCH ARTICLE

## Open Access

# Evidence for acquisition of virulence effectors in pathogenic chytrids

Guiling Sun<sup>†</sup>, Zefeng Yang<sup>†</sup>, Tiffany Kosch, Kyle Summers and Jinling Huang<sup>\*</sup>**Abstract**

**Background:** The decline in amphibian populations across the world is frequently linked to the infection of the chytrid fungus *Batrachochytrium dendrobatidis* (*Bd*). This is particularly perplexing because *Bd* was only recently discovered in 1999 and no chytrid fungus had previously been identified as a vertebrate pathogen.

**Results:** In this study, we show that two large families of known virulence effector genes, crinkler (CRN) proteins and serine peptidases, were acquired by *Bd* from oomycete pathogens and bacteria, respectively. These two families have been duplicated after their acquisition by *Bd*. Additional selection analyses indicate that both families evolved under strong positive selection, suggesting that they are involved in the adaptation of *Bd* to its hosts.

**Conclusions:** We propose that the acquisition of virulence effectors, in combination with habitat disruption and climate change, may have driven the *Bd* epidemics and the decline in amphibian populations. This finding provides a starting point for biochemical investigations of chytridiomycosis.

**Background**

A recent report by IUCN indicates that amphibians are the most endangered group of vertebrates worldwide, with 32 percent listed as threatened with extinction and 159 species already believed to be extinct [1]. Among the several factors often linked to the decline in amphibian populations is the chytrid fungus *Batrachochytrium dendrobatidis* (*Bd*), the causative agent of chytridiomycosis. Ever since the discovery of *Bd* in 1999 [2], numerous cases have been reported for the destruction associated with chytridiomycosis [3,4]. The destructive path of chytridiomycosis fits all of the descriptors of a novel epidemic (i.e. sudden appearance, strong virulence, and rapid transmission) [5]. Investigations of museum specimens collected from as far back as 1902 in Japan [6] and 1938 in Africa [7] have detected *Bd* in skin cross sections. On the other hand, long-term population studies since the 1900's had no records of mass mortalities until the late 1970's when the first *Bd* epidemics were reported [8]. This leads to an interesting question about how pathogenesis suddenly evolved in this previously benign organism.

Many disease organisms existed in non-virulent/com-mensal forms until they acquired some novel traits allowing them to exploit new niches [9]. Among the several possible underlying mechanisms, horizontal gene transfer (HGT) is the most rapid way for an organism to gain novel phenotypes [10] and it has contributed to the origin and spread of pathogenesis in numerous bacteria [9,11] and occasionally in eukaryotes [12,13]. In this study, we show that homologs of two large families of known virulence effector genes were likely acquired by *Bd*. These families have been rapidly duplicated after their acquisition in *Bd* and evolved under strong positive selection. We hypothesize that the acquisition of these virulence effector homologs may partly explain the evolution of chytridiomycosis.

**Results and Discussion**

Phylogenomic analyses identified homologs of two large families of known virulence effectors in *Bd*, including serine peptidases and CRN proteins (Table 1; Additional file 1). Of these, members of the serine peptidase family have been implicated as virulence effectors in *Bd* because of their potential role in degrading host antimicrobial peptides [14]. Previous microarray data also indicated that many serine peptidase genes in *Bd* are highly expressed in the sporangia stage, which is associated

\* Correspondence: [huangj@ecu.edu](mailto:huangj@ecu.edu)

<sup>†</sup> Contributed equally

Department of Biology, East Carolina University, Greenville, NC 27858, USA

**Table 1 Structures and predicted subcellular localization of the CRN protein family in *Bd* JAM 81.**

JGI IDs	Domain structure		Introns	Scaffold location/gene number in the scaffold	ESTs	TargetP	SignalP		iPSORT
	N-	C-					NN	HMM	
26417	typeA	DXX DHA	1	scaffold_10/377		—		—	—
26603	typeA	DXX DHA	2	scaffold_10/377		—	—	—	—
26694		DN17	0	scaffold_10/377		—	—	—	—
90730	typeA	DXX DHA	1	scaffold_10/377		—	—	—	—
90240	typeA	DXX DHA	1	scaffold_9/369		—	—	—	—
26137	typeA	DXX DHA	6	scaffold_9/369		—	—	—	—
26152	typeA	DXX DHA	3	scaffold_9/369		—	—	—	—
26186	typeA	DXX DHA	0	scaffold_9/369		—	—	—	—
90292		DN17	1	scaffold_9/369		—	—	—	—
90343	typeB	DXX DXV	0	scaffold_9/369	6	—	—	—	—
26410	typeA	DXX DHA	5	scaffold_9/369		—	—	—	—
23764	typeA	DXX DHA	2	scaffold_4/596	1	—	—	—	—
34850	typeB	DFB	0	scaffold_4/596	10	—	—	—	—
87524	typeB	DFB	0	scaffold_4/596	12	—	—	—	—
24811	typeA	DXX DHA	2	scaffold_6/516		—	—	—	—
88508	typeA	DXX DHA	0	scaffold_6/516	2	—	—	—	—
25057	typeB	DX8	0	scaffold_6/516	3	—	—	—	—
84882	typeA	DXX DHA	0	scaffold_1/1593	1	—	—	—	—
84908	typeA	DXX DHA	0	scaffold_1/1593		—	—	—	—
34275	typeA	DXX DHA	3	scaffold_1/1593	2	—	—	—	—
22414	typeA	DXX DAB	1	scaffold_1/1593		—	—	—	—
85109	typeA	DXX DHA	0	scaffold_1/1593	1	—	—	—	—
22610	typeA	DXX DHA	6	scaffold_1/1593	1	—	—	—	—
22615	typeA	DXX DHA	1	scaffold_2/629		—	—	—	—
86546	typeB	DFA DDC	0	scaffold_2/629	1	—	—	—	—
86517		DN17	1	scaffold_2/629		—	—	—	—
36642		DBE	0	scaffold_2/629	2	—	S	S	—
23071	typeB	DX8	4	scaffold_2/629	4	—	—	—	—
23077	typeB	DX8	3	scaffold_2/629	3	—	—	—	—
23173	typeA	DXX DAB	3	scaffold_2/629	1	—	—	—	—
23206	typeA	DXX DHA	4	scaffold_3/601		—	—	—	—
23217		DN17	1	scaffold_3/601		—	—	—	—
87221	typeB	DX8	0	scaffold_3/601	2	—	S	—	—
87128		DN17	0	scaffold_3/601		—	—	—	—
23760	typeA	DXX DHA	1	scaffold_3/601	1	—	—	—	—
25252	typeA	DXX DHA	5	scaffold_7/438		—	—	—	—
25352	typeB	DXX DXV	0	scaffold_7/438	2	—	—	—	—
35207		DX8	1	scaffold_7/438	3	—	—	—	—
37012	typeB	DFB	0	scaffold_7/438	10	—	—	—	—
25664	typeA	DXX DXV	1	scaffold_7/438		—	—	—	—
91252	typeA	DXX DHA	1	scaffold_12/268		—	—	—	—
91239		DFB	1	scaffold_12/268	8	—	—	—	—
91238	typeA	DXX DHA	1	scaffold_12/268	1	—	—	—	—
27205	typeA	DXX DHA	2	scaffold_12/268		—	—	—	—
35892	typeB	DFA DDC	0	scaffold_12/268	6	—	—	—	—
91331	typeA	DXX DHA	1	scaffold_12/268		—	—	—	—
89833		DFB	0	scaffold_8/465	8	—	—	—	—
89826	typeA	DXX DHA	1	scaffold_8/465		—	—	—	—
24299	typeA	DXX DHA	3	scaffold_5/527		—	—	—	—

**Table 1 Structures and predicted subcellular localization of the CRN protein family in *Bd* JAM 81. (Continued)**

24333	typeB	DFB	0	scaffold_5/527	14	—	—	—	—
24341	typeA	DXX DAB	2	scaffold_5/527		—	—	—	—
36889	typeB	DFB	0	scaffold_5/527	12	—	—	—	—
87953	typeB	DFB	0	scaffold_5/527	13	—	S	—	—
87954	typeB	DFB	0	scaffold_5/527	14	—	S	—	—
88475	typeA	DXX DHA	0	scaffold_5/527		—	—	—	—
24764	typeB	DFA DDC	0	scaffold_5/527	2	—	—	—	—
88482	typeA	DXX DHA	2	scaffold_5/527	1	—	—	—	—
27291	typeA	DXX DHA	3	scaffold_13/279		—	—	—	—
37407	typeB	DFB	0	scaffold_13/279	14	—	—	—	—
27710	typeA	DXX DHA	1	scaffold_14/258	1	—	—	—	—
27719	typeA	DXX DHA	0	scaffold_14/258		—	—	—	—
36061	typeB	DX8	0	scaffold_14/258	6	—	—	—	—
28349	typeA	DXX DXV	1	scaffold_18/194		—	—	—	—
28355	typeA	DXX DHA	0	scaffold_18/194		M	—	—	—
92602	typeA	DXX DHA	0	scaffold_18/194		—	—	—	—
28463	typeA	DXX DHA	4	scaffold_18/194		—	—	—	—
28466	typeA	DXX DXV	1	scaffold_18/194	1	—	—	—	—
92692	typeB	DFA DDC	1	scaffold_18/194	3	—	—	—	—
28183	typeA	DXX DHA	4	scaffold_17/196		—	—	—	—
28329	typeA	DXX DHA	3	scaffold_17/196		—	—	—	—
28345	typeA	DXX DHA	1	scaffold_17/196		—	—	—	—
28523		DXX DHA	2	scaffold_19/95		S	S	—	—
28594		DN17	3	scaffold_19/95	1	—	—	—	—
28683	typeA	DXX DHA	3	scaffold_20/92		—	—	—	—
28687	typeA	DXX DHA	3	scaffold_20/92		—	—	—	—
26749	typeA	DXX DHA	4	scaffold_11/338		S	S	—	—
90966		DX8	0	scaffold_11/338	1	—	S	—	—
26764	typeA	DXX DAB	3	scaffold_11/338		—	—	—	—
26962	typeB	DX8	0	scaffold_11/338	7	—	S	—	—
26980	typeB	DX8	0	scaffold_11/338	6	—	S	—	—
91063	typeA	DXX DHA	1	scaffold_11/338	1	—	—	—	—
92027	typeA	DXX DHA	1	scaffold_15/227		—	—	—	—
27784	typeA	DXX DAB	2	scaffold_15/227		—	—	—	—
92094	typeA	DXX DHA	1	scaffold_15/227		—	—	—	—

"Type A" and "Type B" show two different types of conserved N-terminal domains of *Bd* CRN proteins, whose detailed information is shown in Additional file 5. "M" and "S" show that the protein is likely localized in mitochondria and the secretory pathway, respectively.

with the infection of keratinized host tissue [14]. Our analyses show that at least 32 serine peptidase genes are present in the *Bd* genome (Additional file 1). Other identifiable homologs of *Bd* serine peptidases are only found in bacteria, where they are more broadly distributed.

The CRN protein family consists of cytoplasmic virulence effectors only known in oomycete plant pathogens [15,16]. These effectors are referred to as "crinkler" proteins because of their association with cell death and leaf crinkling, which parallels the effects of *Bd* on amphibian skin [17]. In oomycete pathogens, the N-terminal region of CRN proteins contains a highly conserved LFLAK

domain that is characteristic of all CRN proteins. The C-terminal region is diverse in domain and motif structure, based on which 24 subfamilies were identified in the oomycete *Phytophthora infestans* [16]. The C-terminal region of CRN proteins also controls virulence and, when expressed in plant cells, may induce cell death [16]. Our analyses identified 84 genes encoding CRN protein homologs in *Bd* (Table 1). These genes are present in genomes of both *Bd* isolates (JAM81 and JEL423) and include eight of the 24 subfamilies of CRN proteins known in oomycetes (Table 2).

Compared to oomycetes, the number of genes in each subfamily is often higher in *Bd*. In particular, the DXX-

**Table 2 Comparison of gene copy numbers of CRN protein subfamilies in *Bd* and oomycetes**

Subfamilies	<i>Bd</i> - JAM81	<i>Bd</i> - JEL423	<i>Pi</i>	<i>Ps</i>	<i>Pc</i>	<i>Pr</i>	<i>Pu</i>	<i>Pb</i> <sup>1</sup>	<i>Pp</i> <sup>1</sup>	<i>Ae</i> <sup>1</sup>
DBE <sup>2</sup>	1	1	1	2	2				y	
DFA-DDC	4	5	1	2	2				y	y
DFB	10	11	8	2	3		1	y		
DXX-DAB	5	5	8	3						
DXX-DHA <sup>2</sup>	44	30	9	1	1					
DXX-DXV <sup>2</sup>	5	5	7	3	1					
DN17	6	6	10	10	4	2		y		y
DX8	9	7	7	4				y		

Abbreviations: *Pi*, *Phytophthora infestans*; *Ps*, *Phytophthora sojae*; *Pc*, *Phytophthora capsici*; *Pr*, *Phytophthora ramorum*; *Pu*, *Pythium ultimum*; *Pb*, *Phytophthora brassicae*; *Pp*, *Phytophthora parasitica*; *Ae*, *Aphanomyces euteiches*.

<sup>1</sup>Sequences were found in EST data.

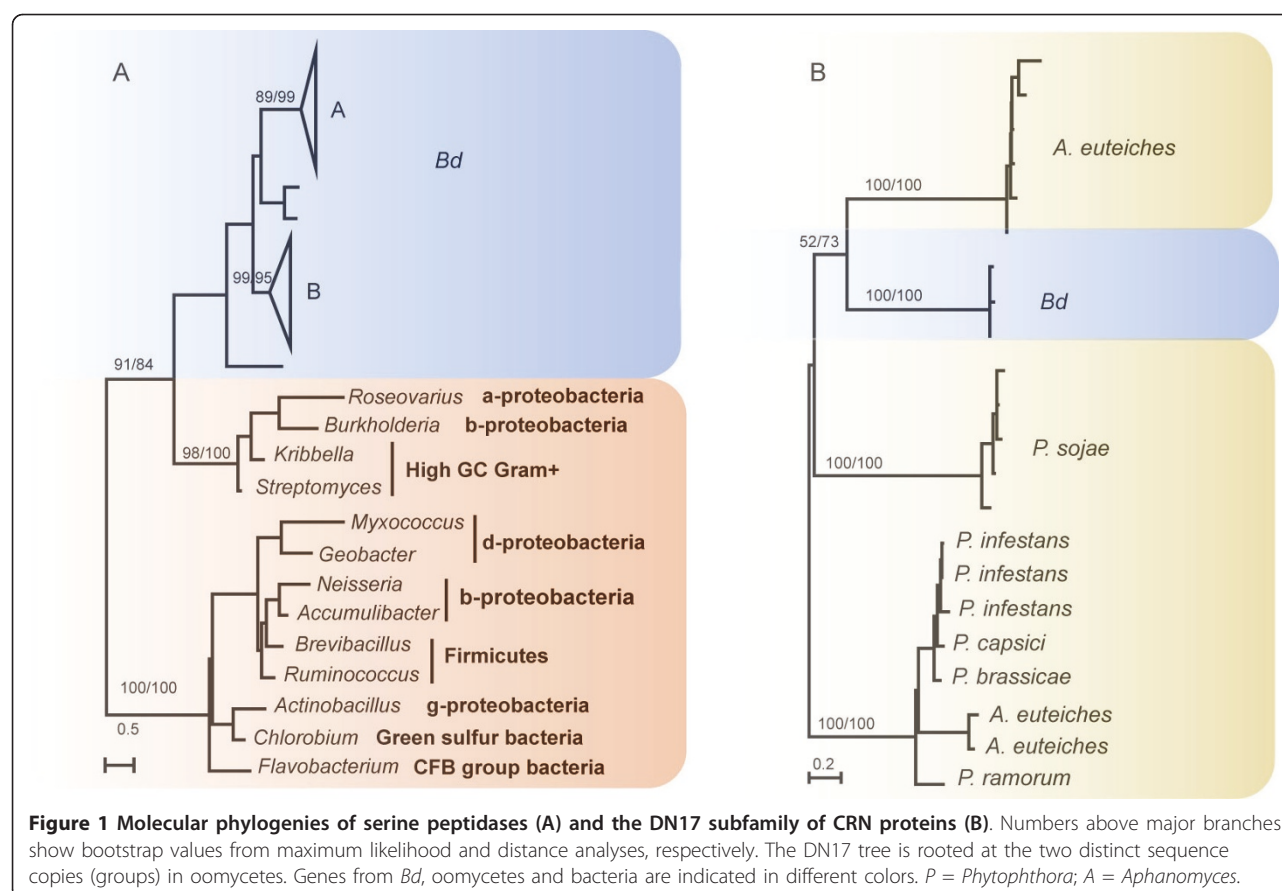
<sup>2</sup>Homologs are present in green plants. All other subfamilies are solely restricted in *Bd* and oomycetes.

DHA subfamily of CRN proteins is highly duplicated in *Bd* and contains 30 and 44 gene copies from the two respective genomes, many of which show very little sequence variation. The homology of CRN proteins from *Bd* and oomycetes is most obvious in the C-terminal effector region, with a sequence identity of up to 46.5%. The N-terminal LFLAK domain characteristic of oomycete CRN proteins also is largely conserved in some identified CRN homologs in *Bd* (Additional file 2). As in oomycete pathogens [16,18], only a fraction of CRN homologs identified in *Bd* are predicted to contain signal peptides (Table 1). However, almost all other identified CRN homologs in *Bd* have short N-terminal extensions of about 20 amino acids in positions corresponding to signal peptides (Table 1 and Additional file 2).

The observation that a gene is uniquely shared by distantly related organisms can be attributed to multiple possible scenarios, in particular differential loss and HGT [19,20]. Although differential loss and HGT can always be invoked as alternative explanations to each other, each scenario becomes increasingly less likely if it requires considerably more corresponding (gene loss or HGT) events [19,21]. In the cases of serine peptidases and CRN proteins, because chytrids and oomycetes belong to opisthokonts and heterokonts respectively, the gene loss scenario entails numerous independent loss events in related taxa. A more parsimonious and plausible scenario is the occurrence of HGT between *Bd* and oomycetes. To further investigate the number and the direction of potential HGT events involved, we performed phylogenetic analyses on serine peptidases and each of the eight CRN subfamilies shared between *Bd* and oomycetes. Our analyses indicate that all serine peptidase homologs in *Bd* form a weakly supported clade (Figure 1A). Given that serine peptidases are more widely distributed in bacteria, it is most likely that serine

peptidases in *Bd* are derived from a single HGT event from bacteria, followed by gene duplication. Analyses of CRN subfamilies show that, in most cases, sequences from *Bd* form a distinct clade whereas those from oomycetes form another, suggesting independent HGT for each CRN subfamily. With paralogous sequence clades as outgroup, our analyses of the DN17 subfamily also show that *Bd* sequences were likely acquired from oomycetes, though the support is still modest (Figure 1B). Such low support is expected given the short length of sequences used in our phylogenetic analyses. Additional evidence consistent with the direction of HGT from oomycetes comes from their higher CRN gene diversity. The fewer CRN subfamilies in *Bd* (8 versus 24 in oomycetes) are in line with the transfer from oomycetes, as we would expect the recipient organism to receive a subset of the gene repertoire of the donor. Nevertheless, for both serine peptidases and the CRN family, we were unable to determine the timing of these HGT events because of the paucity of sequence data from other chytrids and oomycetes.

Because both the serine peptidase and CRN families are highly duplicated in *Bd*, we reasoned that they may confer important functions and, therefore, investigated the possibility that strong selection has been involved in their evolution. The estimate of  $d_N/d_S$  for the serine peptidase family in *Bd* is 0.46, while this value is 0.01 in bacteria. Similarly,  $d_N/d_S$  for each of seven CRN subfamilies (with more than three sequences) is significantly higher in *Bd* than in oomycetes (Table 3). These results indicate that both the serine peptidase and CRN families evolved faster in *Bd* than in oomycetes and bacteria. We further used three likelihood tests (M2a vs. M1a, M8 vs. M7 and M8 vs. M8a) [22] to identify the role of positive selection on the serine peptidase family and seven CRN subfamilies in *Bd* and in oomycetes or bacteria, respectively. *A priori* we stipulated that evidence for positive selection will be sufficient only if all three tests performed show statistical significance. In total, all eight families and subfamilies in *Bd* show strong evidence of positive selection (Table 3; Additional file 3). In contrast, we found that only four subfamilies of CRN proteins (DX8, DXX-DXV, DXX-DHA, and DN17) evolved under positive selection in oomycetes. For the subfamilies that were influenced by positive selection in both *Bd* and oomycetes, the improved branch-site model [23] was used to identify positive selection specifically on the diverging *Bd* lineages. The genes in *Bd* were used as foreground branches and those in oomycetes were used as background. Except for the subfamily DXX-DHA, likelihood tests for the other three CRN subfamilies show statistically significant support for positive selection in the *Bd* branch (Additional file 3). Furthermore, multiple positively selected amino acid sites were



identified using the Bayes Empirical Bayes (BEB) estimation procedure [24], suggesting that positive selection acted specifically on certain amino acid sites of the serine peptidase family and the CRN subfamilies after their acquisition by *Bd*.

Most genes of the serine peptidase family in *Bd* were assigned to two clades in the phylogeny with strong bootstrap support (Figure 1A). We estimated the coefficients of both type I (a site-specific shift in evolutionary rate) [25] and type II functional divergence (a cluster-specific shift of amino acid property) [26] between them. The results showed that only the type I coefficient of functional divergence is significantly greater than zero ( $\theta_1 = 0.5744 \pm 0.0589$ , LRT = 220.3270,  $p < 0.01$ ). These results suggest that site-specific selective constraints have been altered for most of the members of this family, leading to group-specific functional evolution after their diversification. Furthermore, 26 amino acid residues were found to be critical for type I functional divergence with a posterior probability higher than 0.95 (Additional file 3), indicating that the evolutionary rates for these sites have shifted between these two clades. Some amino acid residues in the region of the Peptidase\_S41 domain may also have contributed to the

functional divergence, as there are 11 critical residues that are located in this region.

The acquisition of the serine peptidase and CRN families followed by rapid duplication and positive selection suggests that these acquired genes are likely important in the adaptation of *Bd*. Because both families are known virulence effectors, their presence in *Bd* provides a starting point for biochemical investigations of chytrid pathogenesis. If members of these two families are indeed related to *Bd* virulence, their acquisition may at least partly explain the emergence of chytridiomycosis. Peptidases are ubiquitous tools of bacterial parasites that enable host infection, and have also been identified as key factors in fungal pathogenesis. For example, *Microsporium canis*, a dermatophytic fungal parasite, infects keratinized tissue (e.g. skin and nails) in mammals. This fungus secretes dipeptidyl peptidases, which are homologous to serine peptidases of the S9 family in bacteria [27]. These and other proteases enable the fungus to obtain nutrients from the complex network of insoluble proteins that comprise keratin-rich tissues. As mentioned above, serine peptidases were identified as a class of proteins that are associated with virulence in *Bd* using microarray expression analyses [14]. The CRN



**Table 3 Summary of evolutionary rate and positive selection analyses**

Gene family	Lineage	$d_N/d_S$ <sup>1</sup>	Positive selection <sup>2</sup>	PSS <sup>3</sup>
Serine peptidase	<i>Bd</i>	0.46	+	30, 19*, 14**
	bacteria	0.01	-	-
DXX-DAB	<i>Bd</i>	0.70	+	43, 8*
	oomycetes	0.40	-	-
DFB	<i>Bd</i>	0.64	+	90, 21*, 11**
	oomycetes	0.27	-	-
DX8	<i>Bd</i>	0.88	+	101, 18*, 8**
	oomycetes	0.52	+	41, 12*, 5**
DXX-DXV	<i>Bd</i>	0.65	+	114, 2*
	oomycetes	0.59	+	27, 16*, 12**
DXX-DHA	<i>Bd</i>	0.78	+	61, 18*, 8**
	oomycetes	0.63	+	28, 11*, 4**
DFA-DDC	<i>Bd</i>	0.69	+	36, 17*, 8**
	oomycetes	0.26	-	-
DN17	<i>Bd</i>	0.50	+	25, 8*, 3**
	oomycetes	0.47	+	29, 13*, 9**

<sup>1</sup>The  $d_N/d_S$  ratio average across all sites and lineages under PAML model M0.

<sup>2</sup>“+” signs indicate that positive selection acted on the evolution of these genes, while “-” signs indicate that no positive selection was identified. These results were obtained based on the likelihood ratio tests of three site-specific models (Additional file 1).

<sup>3</sup>PSS indicates positively selected sites ( $\omega > 1$  under PAML model M2a). The number of amino acid residues with the posterior probability (PP) of positive selection greater than 0.05 is shown. \*\*: PP > 0.99; \*: PP > 0.95.

genes were previously only known in oomycetes, which themselves also are important pathogens. Although most attention has been paid to oomycete plant pathogens, there is also a diverse group of oomycetes that infect animals [28]. Some of these pathogens are responsible for epidemics in aquatic animals such as crayfish [28]. The oomycete parasite *Saproglenia* may induce high mortality in salmon populations, both in aquaculture and in the wild [29], with effects similar to *Bd* [30]. Hence, the oomycete lineage is well adapted to parasitize vertebrate hosts. It is not surprising that oomycetes have an arsenal of genes associated with parasitism and that the acquisition of such genes may help convert a commensal chytrid to a parasitic pathogen.

## Conclusions

Our data show that two large families of known virulence effector genes, CRN proteins and serine peptidases, were likely transferred to *Bd* from oomycete pathogens and bacteria, respectively. These acquired virulence effector homologs have been under rapid gene duplication and strong positive selection, suggesting that they may play an important role in the adaptation of *Bd*. Given that pathogenesis is the outcome of microbe-host interactions, the impact of any virulence effector is likely more severe when the host becomes stressed. We

hypothesize that the acquisition of virulence effectors, in combination with habitat disruption and climate change, led to the *Bd* epidemics and the decline in amphibian populations. Further investigations are warranted to fully understand the biological functions of virulence effector homologs identified in our analyses.

## Methods

### Data source and genome screening for HGT candidates

Phylogenomic analyses used a customized database combining sequences from the NCBI non-redundant (nr) database, dbEST, the Taxonomically Broad EST Database (TBestDB) [31] and additional eukaryotic genomes. Annotated protein sequences of the heterokont *Aureococcus anophagefferens*, haptophyte *Emiliania huxleyi*, green algae *Chlorella sp.* and *C. vulgaris*, metazoans *Daphnia pulex*, *Capitella sp.* and *Lottia gigantea* were obtained from the Joint Genome Institute <http://www.jgi.doe.gov>. The annotated genome of the red alga *Cyanidioschyzon merolae* was downloaded from the *Cyanidioschyzon merolae* Genome Project <http://merolae.biol.s.u-tokyo.ac.jp>, and those of fungi *Rhizopus oryzae*, *Allomyces macrogynus*, and *Spizellomyces punctatus* were obtained from their sequencing projects at the Broad Institute of Harvard and MIT <http://www.broadinstitute.org>. Detailed sources for annotated protein sequences and ESTs of two *Bd* isolates and seven oomycetes used in our analyses are shown in Additional file 4. A comprehensive database was created using above data.

Genome screening for horizontally acquired genes was performed using AlienG [32]. AlienG is based on the observation that sequence similarity is correlated to sequence relatedness, where sequence similarity is measured by BLAST bit scores. A query sequence is considered to be a candidate of HGT-derived gene if it is significantly more similar to homologs from a potential donor ( $S_d$ ) than to those from closely related taxa ( $S_r$ ). Specifically, a bit score ratio  $S_d/S_r = 1.5$  was used as threshold to identify candidates of HGT-derived genes in *Bd*. To eliminate potential artifacts, only candidates with expression evidence from ESTs or located in contigs containing more than 5 genes are retained for further analyses.

### Identification of serine peptidase and CRN protein families in *Bd* and oomycetes

Phylogenomic analyses identified homologs of serine peptidases and CRN proteins among the HGT candidates. To retrieve other CRN and serine peptidase homologs in *Bd*, we first downloaded bacterial serine peptidases and annotated oomycete CRNs from both GenBank and oomycete genome projects [16,18]; these downloaded sequences were then used as queries to search (BLASTP) against *Bd* JAM81 annotated protein

sequences. All hits with E-values lower than  $1 \times 10^{-5}$  were retained. To classify the retrieved CRN homologs in *Bd*, we followed the criteria used for the oomycete *P. infestans* [16], which is based on the structures of C-terminal specific domains. As a result, eight subfamilies of CRN homologs were identified in *Bd*. To further estimate the copy numbers for homologs of serine peptidases and each CRN subfamily in *Bd*, we first constructed profiles for the *Bd* serine peptidase domain and the C-terminal specific domain regions of each CRN subfamily using HMMER [33]. The generated profiles were used to search the *Bd* genome, and sequences with E-values lower than  $1 \times 10^{-5}$  were retrieved and further aligned using conserved domain regions as references. Extremely divergent or partial sequences were removed from further analyses. The remaining genes were considered to be putative serine peptidases or CRN proteins. To exclude the possibility of contamination in *Bd* JAM81, the same process was also used to identify genes in the isolate *Bd* JEL423. Because the CRN protein family from four oomycetes (*P. infestans*, *P. sojae*, *P. ramorum*, and *Pythium ultimum*) have been well annotated [16,18], we also used their C-terminal specific domain regions as references to identify CRN proteins in other oomycete datasets, including the *P. capsici* genome and ESTs for three other oomycetes (*P. brassicae*, *P. parasitica*, and *Aphanomyces euteiches*). Partial sequences were excluded and the remaining genes were regarded as CRN proteins. Secretion signals of serine peptidases and CRN proteins were identified using three tools (SignalP [34], TargetP [35], and iPSORT [36]) with default settings. The conserved N-terminal profiles were generated using WebLogo <http://weblogo.berkeley.edu>.

#### Phylogenetic analyses

Protein sequence alignment was carried out using MUSCLE [37] and clustalX [38], followed by visual inspection and manual refinement. Gaps and ambiguously aligned sites were removed manually. The most appropriate protein substitution matrix, rate heterogeneity and invariant sites were determined using ModelGenerator [39] for serine peptidases and each of the eight CRN subfamilies identified in *Bd*. Phylogenetic analyses were performed with a maximum likelihood method using PHYML [40] and a distance method using *neighbor* of PHYLIPNEW v.3.68 [41] in EMBOSS package [42]. Bootstrap support values were estimated using 100 pseudo-replicates. For distance analyses, maximum likelihood distances were calculated using TREE-PUZZLE v.5.2 [43] and PUZZLEBOOT v.1.03 (A. Roger and M. Holder, <http://www.tree-puzzle.de>). The models used in TREE-PUZZLE/PUZZLEBOOT to calculate the maximum likelihood distances were the same as those

determined by ModelGenerator. All other parameters in the analyses used default settings.

#### Detection of positive selection

The  $d_N/d_S$  values for the serine peptidase family and seven CRN subfamilies (with more than three members in *Bd*) were estimated both in *Bd* and in oomycetes or bacteria, respectively. Genes with partial sequences were removed from the analyses. The program PAL2NAL [44] was used to convert protein sequence alignments into the corresponding codon-based DNA alignments. The program CODEML [22] was then used to calculate the  $d_N/d_S$  values.

Maximum likelihood analyses of selective constraints on the serine peptidase family and seven CRN subfamilies were performed using PAML [22] with their phylogenies and the codon alignments as input. For the serine peptidase family and each CRN subfamily, we firstly used the site-specific models [22,45,46] to identify positively selected codon sites in *Bd* and in oomycetes or bacteria, respectively. In this analysis, we compared models M1a vs. M2a, M7 vs. M8 and M8a vs. M8 to identify positive selection. To assess whether positive selection acted on the evolution of each group, the log likelihood values for each pair of nested models were compared using a likelihood ratio test (LRT). In the LRT, twice the log likelihood difference ( $2\Delta$ ) between the two models was compared with the  $\chi^2$  statistics, with degrees of freedom (DF) equal to the difference of the number of parameters. The DFs were 3, 2, and 1 for comparisons M1a vs. M2a, M7 vs. M8 and M8a vs. M8, respectively. If the LRT was statistically significant and  $d_N/d_S$  estimate greater than 1, evidence for positive selection would be considered sufficient for the genes analyzed. Then, the Bayes Empirical Bayes method [24] was used to determine the amino acids most likely responsible for positive selection.

The improved branch-site models [23] were also used to search for positive selection for the serine peptidase family and each CRN subfamily. In these tests, we compared the null model A (model = 2, NSsites = 2, with  $\omega$  fixed to 1) with the alternative model (model = 2, NSsites = 2). For each family or subfamily, the *Bd* branch on the phylogeny was used as foreground, while the oomycetes or bacteria branch used as background. When the LRT suggests positive selection under this model, the codon sites likely evolved under positive selection can be identified through posterior probability from Bayes Empirical Bayes method [24].

#### Analysis of functional divergence

The program DIVERGE [47] was used to estimate the coefficients of both type I ( $\theta_I$ ) and type II ( $\theta_{II}$ ) functional divergence between serine peptidase clades A and

B. Because subfamilies with fewer than four sequences cannot be analyzed using this method, three other serine peptidase proteins were excluded from this analysis. A likelihood ratio test was performed to test the null hypothesis  $\theta = 0$  using  $\chi^2$  distribution with 1 DF. A significant  $\theta_I$  indicates that site-specific altered selection constraints after the divergence of clades, while a significant  $\theta_{II}$  means site-specific shift of amino acid physiochemical property [25,26].

## Additional material

**Additional file 1: Gene structures and subcellular localizations of serine peptidases in *Bd*.** This file contains information about gene identifiers, numbers of exon, and predicted subcellular localizations for serine peptidases that were identified in *Bd* JAM 81.

**Additional file 2: LFLAK domain of the CRN protein family in oomycetes and *Bd*.** This file shows that the LFLAX domain of the CRN protein family is largely conserved between oomycetes and *Bd*.

**Additional file 3: Results of selection analyses.** This file contains results of positive selection analyses using site-specific and branch-site models, respectively. It also shows amino acid residues that are critical to functional divergence between serine peptidase clades A and B.

**Additional file 4: Data sources for *Bd* isolates and oomycetes used in the analyses.** This file shows data types and download sites for two *Bd* isolates and seven oomycetes used in the analyses.

**Additional file 5: Two different conserved N-terminal domains of *Bd* CRN proteins.** This file shows two different types of N-terminal domain for *Bd* CRN proteins. These domain types are also indicated in Table 1.

## Acknowledgements

This work is supported in part by a NSF Assembling the Tree of Life (ATOL) grant (DEB 0830024).

## Authors' contributions

JH designed the study and wrote the manuscript. GS and ZY performed the analyses and participated in manuscript writing. SK and KS wrote the manuscript and contributed to data analyses. All authors read and approved the final manuscript.

## Competing interests

The authors declare that they have no competing interests.

Received: 13 May 2011 Accepted: 8 July 2011 Published: 8 July 2011

## References

- Vié J-C, Hilton-Taylor C, Stuart SN: *Wildlife in a Changing World - An Analysis of the 2008 IUCN Red List of Threatened Species* Gland, Switzerland: UCN; 2009.
- Longcore JE, Pessier AP, Nichols DK: *Batrachochytrium dendrobatidis* gen. et sp. nov., a chytrid pathogenic to amphibians. *Mycologia* 1999, **219**:227.
- Fisher MC, Garner TW, du Preez L, Longcore JE: Global emergence of *Batrachochytrium dendrobatidis* and amphibian chytridiomycosis in space, time, and host. *Annu Rev Microbiol* 2009, **63**:291-310.
- James TY, Litvinseva AP, Vilgalys R, Morgan JA, Taylor JW, Fisher MC, Berger L, Weldon C, du Preez L, Longcore JE: Rapid global expansion of the fungal disease chytridiomycosis into declining and healthy amphibian populations. *PLoS Pathog* 2009, **5**:e1000458.
- Skerratt LF, Berger L, Speare R, Cashings S, McDonald KR, Phillott AD, Hines HB, Kenyon N: Spread of chytridiomycosis has caused the rapid global decline and extinction of frogs. *EcoHealth* 2007, **4**:125-134.
- Goka K, Yokoyama J, Une Y, Kuroki T, Suzuki K, Nakahara M, Kobayashi A, Inaba S, Mizutani T, Hyatt AD: Amphibian chytridiomycosis in Japan: distribution, haplotypes and possible route of entry into Japan. *Mol Ecol* 2009, **18**:4757-4774.
- Weldon C, du Preez LH, Hyatt AD, Muller R, Spears R: Origin of the amphibian chytrid fungus. *Emerg Infect Dis* 2004, **10**:2100-2105.
- Barinaga M: Where have all the froggies gone? *Science* 1990, **247**:1033-1034.
- Ziebuhr W, Ohlsen K, Karch H, Korhonen T, Hacker J: Evolution of bacterial pathogenesis. *Cell Mol Life Sci* 1999, **56**:719-728.
- Gogarten JP, Doolittle WF, Lawrence JG: Prokaryotic evolution in light of gene transfer. *Mol Biol Evol* 2002, **19**:2226-2238.
- Ochman H, Lawrence JG, Groisman EA: Lateral gene transfer and the nature of bacterial innovation. *Nature* 2000, **405**:299-304.
- Ma LJ, van der Does HC, Borkovich KA, Coleman JJ, Daboussi MJ, Di Pietro A, Dufresne M, Freitag M, Grabherr M, Henrissat B, Houterman PM, Kang S, Shim WB, Woloshuk C, Xie X, Xu JR, Antoniw J, Baker SE, Bluhm BH, Breakspear A, Brown DW, Butchko RA, Chapman S, Coulson R, Coutinho PM, Danchin EG, Diener A, Gale LR, Gardiner DM, Goff S, et al: Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* 2006, **446**:367-373.
- Friesen TL, Stukenbrock EH, Liu Z, Meinhardt S, Ling H, Faris JD, Rasmussen JB, Solomon PS, McDonald BA, Oliver RP: Emergence of a new disease as a result of interspecific virulence gene transfer. *Nat Genet* 2006, **38**:953-956.
- Rosenblum EB, Stajich JE, Maddox N, Eisen MB: Global gene expression profiles for life stages of the deadly amphibian pathogen *Batrachochytrium dendrobatidis*. *Proc Natl Acad Sci USA* 2008, **105**:17034-17039.
- Torto TA, Li S, Styer A, Huitema E, Testa A, Gow NA, van West P, Kamoun S: EST mining and functional expression assays identify extracellular effector proteins from the plant pathogen *Phytophthora*. *Genome Res* 2003, **13**:1675-1685.
- Haas BJ, Kamoun S, Zody MC, Jiang RH, Handsaker RE, Cano LM, Grabherr M, Kodira CD, Raffaele S, Torto-Alalibo T, Bozkurt TO, Ah-Fong AM, Alvarado L, Anderson VL, Armstrong MR, Avrova A, Baxter L, Beynon J, Boevink PC, Bollmann SR, Bos JL, Bulone V, Cai G, Cakir C, Carrington JC, Chawner M, Conti L, Costanzo S, Ewan R, Fahlgren N, et al: Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature* 2009, **461**:393-398.
- Berger L, Speare R, Skerratt LF: Distribution of *Batrachochytrium dendrobatidis* and pathology in the skin of green tree frogs *Litoria caerulea* with severe chytridiomycosis. *Dis Aquat Organ* 2005, **68**:65-70.
- Levesque CA, Brouwer H, Cano L, Hamilton JP, Holt C, Huitema E, Raffaele S, Robideau GP, Thines M, Win J, Zerillo MM, Beakes GW, Boore JL, Busam D, Dumas B, Ferriera S, Fuerstenberg SI, Gachon CM, Gaulin E, Govers F, Grenville-Briggs L, Horner N, Hostetler J, Jiang RH, Johnson J, Krajaeun T, Lin H, Meijer HJ, Moore B, Morris P, et al: Genome sequence of the necrotrophic plant pathogen *Pythium ultimum* reveals original pathogenicity mechanisms and effector repertoire. *Genome Biol* 11:R73.
- Huang J, Gogarten JP: Ancient horizontal gene transfer can benefit phylogenetic reconstruction. *Trends Genet* 2006, **22**:361-366.
- Keeling PJ, Palmer JD: Horizontal gene transfer in eukaryotic evolution. *Nat Rev Genet* 2008, **9**:605-618.
- Gogarten JP, Townsend JP: Horizontal gene transfer, genome innovation and evolution. *Nat Rev Microbiol* 2005, **3**:679-687.
- Yang Z: PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 2007, **24**:1586-1591.
- Zhang J, Nielsen R, Yang Z: Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol* 2005, **22**:2472-2479.
- Yang Z, Wong WS, Nielsen R: Bayes empirical bayes inference of amino acid sites under positive selection. *Mol Biol Evol* 2005, **22**:1107-1118.
- Gu X: Functional divergence in protein (family) sequence evolution. *Genetica* 2003, **118**:133-141.
- Gu X: A simple statistical method for estimating type-II (cluster-specific) functional divergence of protein sequences. *Mol Biol Evol* 2006, **23**:1937-1945.
- Vermout S, Baldo A, Tabart J, Losson B, Mignon B: Secreted dipeptidyl peptidases as potential virulence factors for *Microsporium canis*. *FEMS Immunol Med Microbiol* 2008, **54**:299-308.
- Phillips AJ, Anderson VL, Robertson EJ, Secombes CJ, van West P: New insights into animal pathogenic oomycetes. *Trends Microbiol* 2008, **16**:13-19.



29. van West P: *Saprolegnia parasitica*, an oomycete pathogen with a fishy appetite: new challenges for an old problem. *Mycologist* 2006, **20**:99-104.
30. Kiesecker JM, Blaustein AR, Belden LK: **Complex causes of amphibian population declines.** *Nature* 2001, **410**:681-684.
31. O'Brien EA, Koski LB, Zhang Y, Yang L, Wang E, Gray MW, Burger G, Lang BF: **TBestDB: a taxonomically broad database of expressed sequence tags (ESTs).** *Nucleic Acids Res* 2007, **35**:D445-451.
32. Tian J, Sun G, Ding Q, Huang J, Oruganti S, Xie B: **AlienG: an effective computational tool for phylogenetic identification of horizontally transferred genes.** *The third International Conference on Bioinformatics and Computational Biology (BICoB): 23-25 March 2011; New Orleans, Louisiana 2011*, 13-18.
33. Eddy SR: **Profile hidden Markov models.** *Bioinformatics* 1998, **14**:755-763.
34. Bendtsen JD, Nielsen H, von Heijne G, Brunak S: **Improved prediction of signal peptides: SignalP 3.0.** *J Mol Biol* 2004, **340**:783-795.
35. Emanuelsson O, Brunak S, von Heijne G, Nielsen H: **Locating proteins in the cell using TargetP, SignalP and related tools.** *Nat Protoc* 2007, **2**:953-971.
36. Bannai H, Tamada Y, Maruyama O, Nakai K, Miyano S: **Extensive feature detection of N-terminal protein sorting signals.** *Bioinformatics* 2002, **18**:298-305.
37. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** *Nucleic Acids Res* 2004, **32**:1792-1797.
38. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG: **Clustal W and Clustal X version 2.0.** *Bioinformatics* 2007, **23**:2947-2948.
39. Keane TM, Creevey CJ, Pentony MM, Naughton TJ, McLnerney JO: **Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified.** *BMC Evol Biol* 2006, **6**:29.
40. Guindon S, Gascuel O: **A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood.** *Syst Biol* 2003, **52**:696-704.
41. Felsenstein J: **PHYLIP (Phylogeny Inference Package) version 3.61.** Distributed by the author. Book *PHYLIP (Phylogeny Inference Package) version 3.61. Distributed by the author (Editor ed.eds).* City 2005.
42. Rice P, Longden I, Bleasby A: **EMBOSS: the European Molecular Biology Open Software Suite.** *Trends Genet* 2000, **16**:276-277.
43. Schmidt HA, Strimmer K, Vingron M, von Haeseler A: **TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing.** *Bioinformatics* 2002, **18**:502-504.
44. Suyama M, Torrents D, Bork P: **PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments.** *Nucleic Acids Res* 2006, **34**:W609-612.
45. Yang Z, Nielsen R, Goldman N, Pedersen AM: **Codon-substitution models for heterogeneous selection pressure at amino acid sites.** *Genetics* 2000, **155**:431-449.
46. Wong WS, Yang Z, Goldman N, Nielsen R: **Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites.** *Genetics* 2004, **168**:1041-1051.
47. Gu X, Vander Velden K: **DIVERGE: phylogeny-based analysis for functional-structural divergence of a protein family.** *Bioinformatics* 2002, **18**:500-501.

doi:10.1186/1471-2148-11-195

**Cite this article as:** Sun et al.: Evidence for acquisition of virulence effectors in pathogenic chytrids. *BMC Evolutionary Biology* 2011 **11**:195.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

